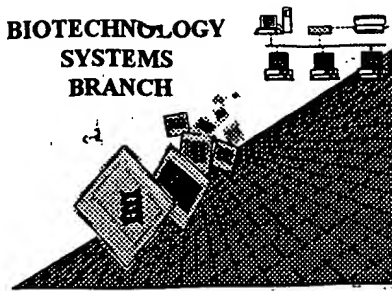


0590
12

BIOTECHNOLOGY
SYSTEMS
BRANCH



**RAW SEQUENCE LISTING
ERROR REPORT**

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/765,061A

Source: OIP

Date Processed by STIC: 8/6/2001

61/2

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: patin3help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW:

Checker Version 3.0

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual Property Organization (WIPO) Standard ST.25.

Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO).

Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the following address:

<http://www.uspto.gov/web/offices/pac/checker>

OIPE

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/765,061A

DATE: 08/06/2001

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

SEQUENCE LISTING

2 (1) GENERAL INFORMATION:
 E--> 3 (iii) NUMBER OF SEQUENCES: 178
 C--> 0 (vi) CURRENT APPLICATION DATA:
 C--> 0 (A) APPLICATION NUMBER: US/09/765,061A
 C--> 0 (B) FILING DATE: 17-Jan-2001

Does Not Comply
 Corrected Diskette Needed
 missing -

ERRORED SEQUENCES

4 (2) INFORMATION FOR SEQ ID NO: 1:
 5 (i) SEQUENCE CHARACTERISTICS:
 6 (A) LENGTH: 6749 bases
 7 (B) TYPE: nucleic acid
 8 (C) STRANDEDNESS: single
 9 (D) TOPOLOGY: linear
 10 (ii) MOLECULE TYPE: DNA (genomic) human
 11 (ix) FEATURE:
 12 (A) NAME/KEY: AIPL1 gene
 13 (B) LOCATION: 17p13.1
 14 (D) OTHER INFORMATION: produces aryl-hydrocarbon
 15 receptor interacting protein-like 1
 16 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 C--> 18 GGCCTCCCAA AGTGCTGGAT TACAGGCGTG AGTCACCGCG CTGGTCCCC TGTCTTCTTT 60
 19 AAGAAAGCTC AGCGGACCTT TTTCCTTCTT GGGGTGGAAC AAAAAGCCAA ATCTAGCACA 120
 20 ACCCTGGGCA GGGGCCCAGA ATCACTGGAA GCAAAGGTGG ATGGGATAGG AGGCGAGGCT 180
 21 GCCTGTGGAC CACTAGGCCG GCCCAGAGTG CTCTGATGAG AAGCCGGGGC GCCTAGGTCA 240
 22 CCGCCCCCAC CGTCTGCCCT TCCCCCACT CCTCTGGCT GGTAAATCC CAGAGTCTCA 300
 23 GCCGCTAAG TGTCTTCCCC GGAGGTGAGA TTATCTCCGC CTGTGCTGGA CACCTCCCTT 360
 24 TCTCTGCAG CCATGGATGC CGCTCTGCTC CTGAACGTGG AAGGGGTCAA GAAAACCATT 420
 25 CTGCACGGGG GCACGGGCGA GCTCCCAAAC TTCATCACCG GATCCCGAGT GAGTGGGGCC 480
 26 CCTCCGAGC AGACAGGGTC CCCCACAGCA GCTTTCAACA TTCCAGGTGT GCCCAAGGC 540
 27 ACTGTAAACA GCTTTCAGCT GTGCCAAAAA AACAGCCAGG CAGCCCCAGC GCTGGGCCTC 600
 28 CGGGGAGCTC CCAGCGTTTA CCCATTGAGG GGGCATTITT GGTACTTTGC AGATTCAACT 660
 29 TTAGCATGGG CTGAGGGGAA GGGCTTTTGG GAATTTTCTG GGGCCCTAAA TGTGAGTGA 720
 30 GAAGAAAGGG AGTCCGAGGA GTCTTGATAT TTGTCCCCAA ATGTCTGTTA GGCTTCCCTG 780
 31 GACTGAAGGG TGCGTCTGTG GCTACAGAAT TCGGGCTTTG GCCAGGCGAG GCGGCTCCCG 840
 32 CCTGTAATCC CAGCACTTTG GGAGGCCAAG ATGGGCAGAT CATGAGGTCA AGAGTTCGAG 900
 33 ACCAGCCTGA CCAACATGTG AAACCCCATC TCTACTGAAA ATACAAAAAT TAGCCAGATG 960
 34 TGCTGTGGCG CCTGTAATCC CAGTTCAGAT ACTCAGGAGA CTTGAGGCAG GAGAATCACT 1020
 35 TGAGCCCAGG AGGTGGAGGT TGCAGTGAGC CGAGATCATA CCACTGCACT CCAACCTGGG 1080
 36 CAACAGAGTG AGACTCTGTC TCAGAAAAAA AAAAAAAAAA AAGAACTCGG GCTTACTTGA 1140
 37 GGAAGGATTT CTGGACGCAC AGGGCTGTGG GGAGTGGAAAT GGGGTCTGTA GGGAGGGGTG 1200
 38 GGTCCCTCCT CCCTAGGGGG TGCAGGCAGG GTGGAGGTGC TCCAGGGGTC TGAGGCATCT 1260
 39 GATGGGGTGA ACTGAGTGAG CTGACCCCTG GGACAGCCCT GGTGTCGGT GGCAAGGGGG 1320
 40 TGGCTTCTGC CGGGCCTTGA ACAGTGTGTC TAGAGCAGAG TGCACCGTCT CGGTGACTAG 1380
 41 GTGATCTTTC ATTTCCGCAC CATGAAATGT GATGAGGAGC GGACAGTCAT TGACGACAGT 1440

all bases must be in
 upper-case letters when using
 old sequence rules format. 8/6/01

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

42 CGGCAGGTGG GCCAGCCCAT GCACATCATC ATCGGAAACA TGTTCAAGCT CGAGGTCTGG 1500
43 GAGATCCTGC TTACCTCCAT GCGGGTGCAC GAGGTGGCCG AGTTCTGGTG CGACACCATC 1560
44 GTAAGTAGGC CCTGCGCGCC TGTCTCCTGG GACTAGTCTT TTCTGGGCTC ACCCACCCTG 1620
45 TTTGCGGGGC TGCTGTGTTT CGGGAAGCT GGGACTCAAG CGAAGCTTTG CAAAGCCAGT 1680
46 CCTGCAAACCT TATTCCCCAC CGTGTGCATG TGAAGATGGA GGAACAAGG GCTGGAAGGG 1740
47 GTGACCCATG CTGTGGCTGG CTGGTGGGGA GCAGGGCTAT GACCAGCAGG AGTGAGCTGG 1800
48 CCCACTTCAC AGTCCTCACA TCTGTGTGTG TGTGTGTGTG TGTGTGTGTG TGTGTGTGTG 1860
49 TGTGTGTGTG AGAGAGAGAG AGAGAGAGAG AGAGAGNNNN NNNNNNTAGC CTTAGGACTT 1920
50 ATTGCAGAGA CCAACACCTA ACAATGTAAT CAGGCAGCCA GTGCAGGACA TAAATAAGTA 1980
51 AGGCAGTGTG CTTTGGGCCA CAAAAGCACG CTCAGCTTGC TGGAAGCCAT GGGTGCCGAG 2040
52 CTGGGGGCTG CTGAGTCAGG GCCAAAGGGG GCCCCTCCCT GCAGTAAGCT GGTTCCTGGG 2100
53 CCTCTCCCTC CTTTGGTCCA GCTCTTAATC CCAACAGGCT CAACAGCCAT CTGCTTGTCT 2160
54 CTTCCATAAA GAGGCAGAAG GCATTTCCGG CTAATCCCGG CCGGTGGGGC GGGCAGGGTG 2220
55 ACCTCTGTCT CTGTGCTGGT GACCTGGAGG CAGAGCTGAA CTGCTGCATA GAGTTTCAGC 2280
56 CCCTTCACTT CACATGTTGC ATGTGGGGCC AGTGCTGGGT CATCTCAGAA GCCGGTCCAA 2340
57 GGAGATGGGT TCTCAGGGAG CCTAGTTGGG GAAACTGAGG CCCAGCATAC ATACAGCAGG 2400
58 CCTCGCTGAG GCCGCACGGC GGATCTTCCC AGCCCTCCTT CATCCCAAGG GTGGCAAAC 2460
59 CAGCTCCCAT GCTGGCTGAA GCTGTGATGA GCCAGATCTA TATCTGCACC ATCTCATTTA 2520
60 ATCCCTACAG CAGCCCTAAT ATCGAACAGG AGCAACCCAG GGAAGTGAAG TTCAGAGAAG 2580
61 TGCAGAGACC TGGGCTCACC GCTAACCTGC AGCACTGCCA GGACACCAA GCGACTCTCT 2640
62 TGGACCTGG AGTCTGCTC CTCTACTGTC CCCACTGTC CCTTCTGCG AGTCATAGG 2700
63 TTTGCAGAGG TCAGGGTTTC CCTGGGGCAG AGATGTGTTA CAGTGGACCA CAAGGGCCAG 2760
64 AAGAGGCAGC CGGAGGCTAA CAGCATATGG CCTCTGGAGC CAGGTTTGAA TCCTGGCTGC 2820
65 GTCATTTCTT AGCTGTGTGA CCTTAAGCAA GTTGCTTGCG TCTCTGGGCT GTAGTTTCCC 2880
66 CATCCGTAAA ATGGGATAAT AGTGCCTGCC TTGAATTGTC ATAAGGATTG AAGGGGCTCA 2940
67 TAACAGTGTG AAGTGCTTTG CCTGGCACAC AGTTAACCAC AGTTAGTATG AGTGGCATAG 3000
68 TGAGGGAGCA GGATTCCTCC CAGGAGGGGC TCTGAGTGA GGCCTTTTAT GGCCACCTA 3060
69 GCTCTGGGCA GGTAGCCTGG ATGCCATCCA TCCGTTTATC CCCACAGCAC ACGGGGTCT 3120
70 ACCCATCCT RTCCCGGAGC CTGAGGCAGA TGGCCAGGG CAAGGACCC ACAGAGTGGC 3180
71 ACGTGACAC GTGCGGGCTG GCCAACATGT TCGCCTACCA CACGCTGGGC TACGAGGACC 3240
72 TGGACGAGCT GCAGAAGGAG CCTCAGCCTC TGGTCTTTGT GATCGAGCTG CTGCAGGTGG 3300
73 GGCTGGGGTT GGCAGGGCTG GAGGGCTGTG CCAGCACTGG AGAGGGACAG CGGGCATCAT 3360
74 GGGCACCCCC ACCCACTGG CCACTGGACA GTGCCCTGTT TCTGTTTAGA TAATACGAGA 3420
75 GGGTTCATAA GCCATGGGAG AATACGAATT TGAACAAAAA GTCCTCTGAT TTTCCACAA 3480
76 GAAAAGTCCT TTGGTGCTGG GCATGGTGGC CCACGCCTGT AATCCTAGCA CTTTGGGAGG 3540
77 CCGAGGGGGT TGGATCACCT GAGGTCAGGA GTTCGAAGAC CAGCCTGGCC AACATGGTAA 3600
78 AACCCGTCT CTATTAAAA CACAAAAATT AACCAGGTGT GGTGGTGCAT GCCTGTAATC 3660
79 AATCCAGCT ACTTGGGAAT TTGAGGCATG AGAATTGCTT GAACCTGGA GTGGAGGTTG 3720
80 CAGTGAGCAG AGATCATGTC AGTGCATTTT AACCTGGGTG ACAGAGTGAG ACTCCATGTC 3780
81 CAAAAAAAAG AAAAAAAA AAAGTCCACT TGGAACCAGT TTTTAAAAAT GTGATTCATT 3840
82 TTCATTGTGG AGGCATTTTA TCCACTTCCA CTTTCATTTT CAGGAGTTGG AGATTATAAC 3900
83 CGCCTCCTTG GTTCTGTGG TTTGTGGGTT CAGACTTGGT TCTCTNGTGG CGGGAGAGGC 3960
84 TGCATGGAAC TCCCCACATC CTCCCAACCA GGAGCCCCAG AGTGATTGGC AGCGCGTGTT 4020
85 TGTGATTGG TGAGAGAGGG TTAGGGCCAG GGTCAAGGTC AGGTCAGGAC TCAGCTTATG 4080
86 GCCAAGACTG AGGCTCAGCC TGAGAGCTAT GTGGGTGAAT AAAATAAAAT AAGAACTGTG 4140
87 TCAACCAAGG GCCCCTTACA GGCTTGCTGT CACAGTTGTG TGGTCTGTGC ACTGCACAAG 4200
88 GTGCACCGGC ATCTCCTCCA AGGTGCTCAT TATAGACATT GTATATTGGT ATTTCCATAA 4260
89 TGAGAAGTTT CCAGCAGATG GCAATAGTGT ATTGTTCTAA CAAAACGAGT ATTCGTGACA 4320
90 ATTTTCTGAA TATTAGAAGT GAAGTGTCTT GATGAACGGG CACCTTTTCC TAGTTTGCAC 4380

```

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

91 AAAGACATTG ATTTAGGGCA GGGTTTTTCGG CGTTGTTGCT TCTTTCCTT GTCTGTATGC 4440
92 ACTTGACCAG CAAGCATGAC TTCAGGGAGA TGTGCCACAG GGTCTGTTT TTCGGGTCTC 4500
93 TGATGGGGTG CAGGCCCCTG GGGTCCCCTG CTCACTGACC TGCAGCTCTG GGGCCAGGTT 4560
94 GATGCCCCGA GTGATTACCA GAGGGAGACC TGGAACTGA GCAATCATGA GAAGATGAAG 4620
95 GCGGTGCCCG TCCTCCACGG AGAGGGAAAT CGGCTCTTCA AGCTGGGCCG CTACGAGGAG 4680
96 GCCTCTTCCA AGTACCAGGA GGCCATCATC TGCCTAAGGA ACCTGCAGAC CAAGGTCAGA 4740
97 GGCCGCTGGC CAGGGGTGGG AAGTGGCGCT GACTCTGGGG GGCCTGCCCC GTGCCGGCCA 4800
98 GGGTGGGGCG GGGGTGGGC AGCTGCCTGA GGTTCATGGT GACCTTCTCC CTGGGCAGGA 4860
99 GAAGCCATGG GAGGTGCAGT GGCTGAAGCT GGAGAAGATG ATCAATACTC TGATCCTCAA 4920
100 CTACTGCCAG TGCCTGCTGA AGAAGGAGGA GTACTATGAG GTGCTGGAGC ACACCAGTGA 4980
101 TATTCTCCGG CACCACCCAG GTGCGCGGGG CTGCAGGGGC GGACAGTGAG GGGGCGCCCA 5040
102 GCCCAGGGCC ACGGAGACAC CTGCCATAGC CTTCTTGGAC TTTTCTTTCC ACCCCACCAG 5100
103 GGCACCAAAC CTTGTCTCCA CCCAGCCGGG TTTCCCCGAG TGTGTAAGT AATTGTGGGT 5160
104 GATGGATGGG CAGTGCTTGG CGCGGGGCGG CCTTTATTTT AATGTGTGTT TGAACACTTA 5220
105 CCCAGGAAGC TCGCCAAGCT TGTGATTTCA GCGGAACGGT AAACAGGCGT TTAAAAAGAG 5280
106 GGGCAATCAA TATAGGGAAA AATATTATGA TGTCGGTACT AGTACTGGTG TTGCGAGGAT 5340
107 ATGGCACCGC AGTACTAGAT TGAATTAATG CTCGAATCGT GCTCACAGTA AAAACATCCA 5400
108 GCCCTGGCT CATGCATCAG GCACACGTCG TCTGCGTTTA TTATCTCATT TAATCCTCAT 5460
109 AATCCTCATA ATCACCATAT GAGGGAGGTG CAGGGAAAGG GGCCTGAAGG TTATCTAATT 5520
110 TAGGTAGCGT CTATAAGAAA AATAAAACAA AGTTATGAAT ATAAAATTAC TCACAGGGCC 5580
111 TTAAAAAGGA GAGGAGGAGG TACTGCTATT ATGATCATCA TCTCCATCTT ACAGTTGAGG 5640
112 AAACCGAGGG ATGGGGGATA CAGAGAGGTT AAGGATCATG GCGGGGCTGA GGGTCTTGGA 5700
113 GGCTGGTGAG TCCAGCTGG GCTGGGGCTG CCTCTGAGGC TGGGAAGGGA GCTGTAGCTG 5760
114 GATGCTCCCT GCTCCCCACA GGCATCGTGA AGGCCTACTA CGTGCCTGCC CGGGCTCACG 5820
115 CAGAGGTGTG GAATGAGGCC GAGGCCAAGG CGGACCTCCA GAAAGTGCTG GAGCTGGAGC 5880
116 CGTCCATGCA GAAGGCGGTG CGCAGGGAGC TTGAGGCTGC TGGAGAACCG CATGGCGGAG 5940
117 AACAGGAGGA GGAGCGGCTG CGCTGCCGGA ACATGCTGAG CCAGGGTGCC ACGCAGCCTC 6000
118 CCGCAGAGCC ACCCAGAGAG CCACCCGCAC AGTCATCCAC AGAGCCACCT GCAGAGCCAC 6060
119 CCACAGCACC ATCTGCAGAG CTGTCCGCAG GGCCCCCTGC AGAGCCAGCC ACAGAGCCAC 6120
120 CCCCCTCCCC AGGGCACTCG CTGCAGCACT GAGCCCCCTG AGGCCACAG CCACCCAGGC 6180
121 AGGGAGCAAG TGGCCTGGTC ACTTCTGGTT CGATTGACCA GGATCGTGGT GTCACTTTTT 6240
122 AAAATTTAAA ATTAATTTTT GAAATCAAAG TCAGACACAC CCATGGTAAA AAAAAAAAAA 6300
123 AAAACAATCC CAAGGGTACA GAAGAGCTTA TGAATAAAG TAGTTTTCTC CTCTACCCCT 6360
124 CTCATTCCTT CCGTGCCATG GTTTTAATTG ACCCTGTTTT TAATCTTCT GGTAGTTTTC 6420
125 TCTATTCCA AGTAATCTGT TTAAATCAGT TTCTAGATTT TACCCCATGT CAATGACAAA 6480
126 TGAGGATTTG ATGCTCTGAT CCTTTCTCAT GCCTGATACC CCTCCCTGTC TCCCCATTTT 6540
127 GGATAGTTAC ATTTGGGGGT CATCTCGGTG ATTTTGTAA CTTTACGCAG GACACTTAGA 6600
128 GCTCTCTAGA ATCCCACTGA CTTTAGTGGG GTCTTGATGT AGGGTGGGCA AGCCCCGACA 6660
E--> 129 CTGGAGCTTA GCCTGAGAGG GGTTCCTGC

```

(6749) 6689

333 (2) INFORMATION FOR SEQ ID NO: 8:

334 (i) SEQUENCE CHARACTERISTICS:

335 (A) LENGTH: 1129 bases

336 (B) TYPE: nucleic acid

337 (C) STRANDEDNESS: single

338 (D) TOPOLOGY: linear

339 (ii) MOLECULE TYPE: cDNA Squirrel monkey

340 (ix) FEATURE:

341 (A) NAME/KEY: AIPL1 gene

342 (B) LOCATION:

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

343          (D) OTHER INFORMATION: produces aryl-hydrocarbon
344 receptor interacting protein-like 1
345          (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
C--> 347 ATGGATGCCG CTCTGCTCCT GAACGTGGAA GGGGTCAAGA AGACCATTCT GCACGGGGGC 60
348 ACGGGCGAGC TCCCAAATTT CATCACCGGA TCCCGAGTGA TCTTTCATTT CCGCACCATG 120
349 AAATGTGATG AGGAGCGGAC GGTGATTGAC GACAGCAGGG AGGTGGGCCA GCCCATGCAC 180
350 ATCATCATCG GGAACATGTT CAAGCTGGAG GTCTGGGAGA TCCTGCTCAC GTCCATGCGG 240
351 GTGCGAGAGG TGGCCGAGTT CTGGTGCAC ACCATCCACA CGGGGGTCTA CCCCATCCTG 300
352 TCCCGGAGCC TGCAGCAGAT GGCCAGGGC AAGGACCCGA CGGAGTGGCA TGTGCACACG 360
353 TCGGGGCTGG CCAACATGTT CGCCTACCAC ACGCTGGGCT ACGAGGACCT GGATGAGCTG 420
354 CAGAAGGAGC CTCAGCCTCT GATCTTTGTG ATCGAGCTGC TGCAGGTTGA TGCCCCAAGT 480
355 GATTACCAGA GGGAGACCTG GAACCTGAGC AATCACGAGA AGATGAAGGT GGTGCCCCGTC 540
356 CTCCATGGAG AAGGAAATAG GCTCTTCAAG CTGGGCCGCT ACGAGGAGGC CTCTTCCAAG 600
357 TACCAGGAGG CCATCATCTG CCTAAGGAAC CTGCAGACCA AGGAGAAACC CTGGGAGGTG 660
358 CAGTGGCTGA AGCTGGAGAA GATGATCAAT ACCCTGATCC TCAACTACTG TCAGTGTCTG 720
359 CTGAAGAAGG AGGAGTACTA CGAGGTCCTG GAGCATACCA GTGACATTCT CCGGCACCAC 780
360 CCAGGCATTG TGAAGGCCTA CTATGTGCGC GCCCGGGCTC ACGCGGAGGT GTGGAACGAG 840
361 GCCGAGGCCA AGGCGGACCT CCAGAAAGTG CTGGAGCTGG AGCCGTCCAT GCAGAAGGCG 900
362 GTGCGCAGG AGCTGAGGCT GCTGGAGAAG CGCATGGCGG AGAAGCAGGA GGAGGAGCGG 960
363 CTGCGCTGCC GCAACATGCT CAGCCAGGGG GCCACGTGGT CCCCCGCGGA GCCACCCGCA 1020
364 GAGCCACCTG CAGAGTCATC CACAGAGCCA CCGCAGAGC CACCTGCAGA GCCACCTGCA 1080
E--> 365 GAGCTAACCT TGACCCCGGG GCACCCACTA CAGCACTGA 1129 11/9
383 (2) INFORMATION FOR SEQ ID NO: 10:
384 (i) SEQUENCE CHARACTERISTICS:
385 (A) LENGTH: 15 bases
386 (B) TYPE: nucleic acid
387 (C) STRANDEDNESS: single
388 (D) TOPOLOGY: linear
389 (ii) MOLECULE TYPE: DNA (genomic)
390 (ix) FEATURE:
391 (A) NAME/KEY: AIPL1 Trp88X mutation
392 (B) LOCATION: 86...90
393 (D) OTHER INFORMATION:
394 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
C--> 396 GAG TTC TGA TGC GAC 15
E--> 397 X use Xaa
447 (2) INFORMATION FOR SEQ ID NO: 14:
448 (i) SEQUENCE CHARACTERISTICS:
449 (A) LENGTH: 15 bases
450 (B) TYPE: nucleic acid
451 (C) STRANDEDNESS: single
452 (D) TOPOLOGY: linear
453 (ii) MOLECULE TYPE: DNA (genomic)
454 (ix) FEATURE:
455 (A) NAME/KEY: AIPL1 Gln163X mutation
456 (B) LOCATION: 161...165
457 (D) OTHER INFORMATION:
458 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:
C--> 460 GAT TAC TAG AGG GAG 15

```

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

E--> 461
479 (2) INFORMATION FOR SEQ ID NO: 16:
480 (i) SEQUENCE CHARACTERISTICS:
481 (A) LENGTH: 15 bases
482 (B) TYPE: nucleic acid
483 (C) STRANDEDNESS: single
484 (D) TOPOLOGY: linear
485 (ii) MOLECULE TYPE: DNA (genomic)
486 (ix) FEATURE:
487 (A) NAME/KEY: AIPL1 Trp278X mutation
488 (B) LOCATION: 276...280
489 (D) OTHER INFORMATION:
490 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:
C--> 492 GAG GTG TGA AAT GAG 15
E--> 493
495 (2) INFORMATION FOR SEQ ID NO: 17:
496 (i) SEQUENCE CHARACTERISTICS:
497 (A) LENGTH: 15 bases
498 (B) TYPE: nucleic acid
499 (C) STRANDEDNESS: single
500 (D) TOPOLOGY: linear
501 (ii) MOLECULE TYPE: DNA (genomic)
502 (ix) FEATURE:
503 (A) NAME/KEY: AIPL1 IVS2-2A to G mutation
504 (B) LOCATION:
505 (D) OTHER INFORMATION:
506 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:
C--> 508 TCC CCA CGG CAC ACG 15
E--> 509 IVS2-SA-> G delete
543 (2) INFORMATION FOR SEQ ID NO: 20:
544 (i) SEQUENCE CHARACTERISTICS:
545 (A) LENGTH: 13 bases
546 (B) TYPE: nucleic acid
547 (C) STRANDEDNESS: single
548 (D) TOPOLOGY: linear
549 (ii) MOLECULE TYPE: DNA (genomic)
550 (ix) FEATURE:
551 (A) NAME/KEY: AIPL1 Pro351del12 mutation
552 (B) LOCATION: Pro351
553 (D) OTHER INFORMATION: TGCAGAGCCACC deleted
554 sequence
555 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:
C--> 557 G CCA CCC ACA GCA 13
E--> 558 del TGCAGAGCCACC delete
576 (2) INFORMATION FOR SEQ ID NO: 22:
577 (i) SEQUENCE CHARACTERISTICS:
578 (A) LENGTH: 13 bases
579 (B) TYPE: nucleic acid
580 (C) STRANDEDNESS: single

```

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

581         (D) TOPOLOGY: linear
582     (ii) MOLECULE TYPE: DNA (genomic)
583     (ix) FEATURE:
584         (A) NAME/KEY: AIPL1 Ala336del2 mutation
585         (B) LOCATION: Ala336 2 base deletion
586         (D) OTHER INFORMATION: AG deleted sequence
587     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:
C--> 589         T CCC GCA GCC ACC                                     13
E--> 590         del AG delete
592 (2) INFORMATION FOR SEQ ID NO: 23:
593     (i) SEQUENCE CHARACTERISTICS:
594         (A) LENGTH: 15 bases
595         (B) TYPE: nucleic acid
596         (C) STRANDEDNESS: single
597         (D) TOPOLOGY: linear
598     (ii) MOLECULE TYPE: DNA (genomic)
599     (ix) FEATURE:
600         (A) NAME/KEY: AIPL1 Cys42X mutation
601         (B) LOCATION: 40...44
602         (D) OTHER INFORMATION:
603     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:
C--> 605         ATG AAA TGA GAT GAG                                     15
E--> 606         XAA
608 (2) INFORMATION FOR SEQ ID NO: 24:
609     (i) SEQUENCE CHARACTERISTICS:
610         (A) LENGTH: 12 bases
611         (B) TYPE: nucleic acid
612         (C) STRANDEDNESS: single
613         (D) TOPOLOGY: linear
614     (ii) MOLECULE TYPE: DNA (genomic)
615     (ix) FEATURE:
616         (A) NAME/KEY: AIPL1 Leu257del9 mutation
617         (B) LOCATION: Leu 257 9 base deletion
618         (D) OTHER INFORMATION: CTCCGGCAC deleted sequence
619     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:
C--> 621         GAT ATT CAC CCA                                     12
E--> 622         del CTCCGGCAC
624 (2) INFORMATION FOR SEQ ID NO: 25:
625     (i) SEQUENCE CHARACTERISTICS:
626         (A) LENGTH: 21 bases
627         (B) TYPE: nucleic acid
628         (C) STRANDEDNESS: single
629         (D) TOPOLOGY: linear
630     (ii) MOLECULE TYPE: DNA (genomic)
631     (ix) FEATURE:
632         (A) NAME/KEY: AIPL1 Val33ins8 mutation
633         (B) LOCATION: Val 33 8 base insertion
634         (D) OTHER INFORMATION: GTGATCTT inserted sequence
635     (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

```

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

C--> 637          GAC TAG GTG ATC TTG TGA TCT          21
E--> 638          ins GTGATCTT delete
640 (2) INFORMATION FOR SEQ ID NO: 26:
641 (i) SEQUENCE CHARACTERISTICS:
642 (A) LENGTH: 12 bases
643 (B) TYPE: nucleic acid
644 (C) STRANDEDNESS: single
645 (D) TOPOLOGY: linear
646 (ii) MOLECULE TYPE: DNA (genomic)
647 (ix) FEATURE:
648 (A) NAME/KEY: AIPL1 IVS1-9G to A Benign
649 Variants/Polymorphisms
650 (B) LOCATION:
651 (D) OTHER INFORMATION:
652 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:
C--> 654          CTC AGT GAC TAG          12
E--> 655          G--A
657 (2) INFORMATION FOR SEQ ID NO: 27:
658 (i) SEQUENCE CHARACTERISTICS:
659 (A) LENGTH: 12 bases
660 (B) TYPE: nucleic acid
661 (C) STRANDEDNESS: single
662 (D) TOPOLOGY: linear
663 (ii) MOLECULE TYPE: DNA (genomic)
664 (ix) FEATURE:
665 (A) NAME/KEY: AIPL1 IVS2+66G to C Benign
666 Variants/Polymorphisms
667 (B) LOCATION:
668 (D) OTHER INFORMATION:
669 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:
C--> 671          TTT GCC GGG CTG          12
E--> 672          G--C
674 (2) INFORMATION FOR SEQ ID NO: 28:
675 (i) SEQUENCE CHARACTERISTICS:
676 (A) LENGTH: 12 bases
677 (B) TYPE: nucleic acid
678 (C) STRANDEDNESS: single
679 (D) TOPOLOGY: linear
680 (ii) MOLECULE TYPE: DNA (genomic)
681 (ix) FEATURE:
682 (A) NAME/KEY: AIPL1 IVS2-88C to T Benign
683 Variants/Polymorphisms
684 (B) LOCATION:
685 (D) OTHER INFORMATION:
686 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:
C--> 688          TCC TCT CAG GAG          12
E--> 689          C--T
691 (2) INFORMATION FOR SEQ ID NO: 29:
692 (i) SEQUENCE CHARACTERISTICS:

```


RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

693      (A) LENGTH: 12 bases
694      (B) TYPE: nucleic acid
695      (C) STRANDEDNESS: single
696      (D) TOPOLOGY: linear
697      (ii) MOLECULE TYPE: DNA (genomic)
698      (ix) FEATURE:
699          (A) NAME/KEY: AIPL1 IVS2-14G to A Benign
700 Variants/Polymorphisms
701      (B) LOCATION:
702      (D) OTHER INFORMATION:
703      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:
C--> 705          ATC CAT TTA TCC                      12
E--> 706            G  A  
708 (2) INFORMATION FOR SEQ ID NO: 30:
709      (i) SEQUENCE CHARACTERISTICS:
710          (A) LENGTH: 12 bases
711          (B) TYPE: nucleic acid
712          (C) STRANDEDNESS: single
713          (D) TOPOLOGY: linear
714      (ii) MOLECULE TYPE: DNA (genomic)
715      (ix) FEATURE:
716          (A) NAME/KEY: AIPL1 IVS2-10A to C Benign
717 Variants/Polymorphisms
718      (B) LOCATION:
719      (D) OTHER INFORMATION:
720      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:
C--> 722          CGT TTC TCC CCA                      12
E--> 723            A  C  
725 (2) INFORMATION FOR SEQ ID NO: 31:
726      (i) SEQUENCE CHARACTERISTICS:
727          (A) LENGTH: 12 bases
728          (B) TYPE: nucleic acid
729          (C) STRANDEDNESS: single
730          (D) TOPOLOGY: linear
731      (ii) MOLECULE TYPE: DNA (genomic)
732      (ix) FEATURE:
733          (A) NAME/KEY: AIPL1 IVS3-25T to C Benign
734 Variants/Polymorphisms
735      (B) LOCATION:
736      (D) OTHER INFORMATION:
737      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:
C--> 739          CTG CCC CAC TGA                      12
E--> 740            T  C  
742 (2) INFORMATION FOR SEQ ID NO: 32:
743      (i) SEQUENCE CHARACTERISTICS:
744          (A) LENGTH: 12 bases
745          (B) TYPE: nucleic acid
746          (C) STRANDEDNESS: single
747          (D) TOPOLOGY: linear

```

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

748      (ii) MOLECULE TYPE: DNA (genomic)
749      (ix) FEATURE:
750          (A) NAME/KEY: AIPL1 IVS3-21T to C Benign
751 Variants/Polymorphisms
752          (B) LOCATION:
753          (D) OTHER INFORMATION:
754      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:
C--> 756          CCT CAC CGA CCT                      12
E--> 757          TTC
759 (2) INFORMATION FOR SEQ ID NO: 33:
760      (i) SEQUENCE CHARACTERISTICS:
761          (A) LENGTH: 12 bases
762          (B) TYPE: nucleic acid
763          (C) STRANDEDNESS: single
764          (D) TOPOLOGY: linear
765      (ii) MOLECULE TYPE: DNA (genomic)
766      (ix) FEATURE:
767          (A) NAME/KEY: AIPL1 IVS5+18G to A Benign
768 Variants/Polymorphisms
769          (B) LOCATION:
770          (D) OTHER INFORMATION:
771      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:
C--> 773          AGG AGC GGA CAG                      12
E--> 774          GAA
912 (2) INFORMATION FOR SEQ ID NO: 42:
913      (i) SEQUENCE CHARACTERISTICS:
914          (A) LENGTH: 20 bases
915          (B) TYPE: nucleic acid
916          (C) STRANDEDNESS: single
917          (D) TOPOLOGY: linear
W--> 918      (ii) MOLECULE TYPE: DNA Primer
919      (ix) FEATURE:
920          (A) NAME/KEY: AIPL1 primer
921          (B) LOCATION:
922          (D) OTHER INFORMATION: 5' to 3' order
923      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:
E--> 925      5'-AAGAAAACCATTCCTGCACGG-3'
927 (2) INFORMATION FOR SEQ ID NO: 43:
928      (i) SEQUENCE CHARACTERISTICS:
929          (A) LENGTH: 19 bases
930          (B) TYPE: nucleic acid
931          (C) STRANDEDNESS: single
932          (D) TOPOLOGY: linear
W--> 933      (ii) MOLECULE TYPE: DNA Primer
934      (ix) FEATURE:
935          (A) NAME/KEY: AIPL1 primer
936          (B) LOCATION:
937          (D) OTHER INFORMATION:
938      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

```

Per sequence rules,
group all
non-coding
bases into
10's, with a
space between
groups

delete - do not
show
primer markers

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

E--> 940 5'-TGCAGCTCGTCCAGGTCCT-3' *same* 19
 942 (2) INFORMATION FOR SEQ ID NO: 44:
 943 (i) SEQUENCE CHARACTERISTICS:
 944 (A) LENGTH: 17 bases
 945 (B) TYPE: nucleic acid
 946 (C) STRANDEDNESS: single
 947 (D) TOPOLOGY: linear
 W--> 948 (ii) MOLECULE TYPE: Primer DNA
 949 (ix) FEATURE:
 950 (A) NAME/KEY: AIPL1 primer
 951 (B) LOCATION:
 952 (D) OTHER INFORMATION:
 953 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:
 E--> 955 5'-GACACCTCCCTTTCTCC-3' *same* 17
 957 (2) INFORMATION FOR SEQ ID NO: 45:
 958 (i) SEQUENCE CHARACTERISTICS:
 959 (A) LENGTH: 18 bases
 960 (B) TYPE: nucleic acid
 961 (C) STRANDEDNESS: single
 962 (D) TOPOLOGY: linear
 963 (ii) MOLECULE TYPE: Primer DNA (genomic) human
 964 (ix) FEATURE:
 965 (A) NAME/KEY: AIPL1 primer
 966 (B) LOCATION:
 967 (D) OTHER INFORMATION:
 968 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:
 E--> 970 5'-GCTGGGGCTGCCTGGCTG-3' 18
 972 (2) INFORMATION FOR SEQ ID NO: 46:
 973 (i) SEQUENCE CHARACTERISTICS:
 974 (A) LENGTH: 20 bases
 975 (B) TYPE: nucleic acid
 976 (C) STRANDEDNESS: single
 977 (D) TOPOLOGY: linear
 978 (ii) MOLECULE TYPE: Primer DNA (genomic) human
 979 (ix) FEATURE:
 980 (A) NAME/KEY: AIPL1 Primer
 981 (B) LOCATION:
 982 (D) OTHER INFORMATION:
 983 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:
 E--> 985 5'-CCGAGTGATTACAGAGGGA-3' 20
 987 (2) INFORMATION FOR SEQ ID NO: 47:
 988 (i) SEQUENCE CHARACTERISTICS:
 989 (A) LENGTH: 20 bases
 990 (B) TYPE: nucleic acid
 991 (C) STRANDEDNESS: single
 992 (D) TOPOLOGY: linear
 993 (ii) MOLECULE TYPE: Primer DNA (genomic) human
 994 (ix) FEATURE:
 995 (A) NAME/KEY: AIPL1 Primer

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

996         (B) LOCATION:
997         (D) OTHER INFORMATION:
998         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:
E--> 1000         5'-TGAGCTCCAGCACCTCATAG-3'                20
1002 (2) INFORMATION FOR SEQ ID NO: 48:
1003         (i) SEQUENCE CHARACTERISTICS:
1004             (A) LENGTH: 18 bases
1005             (B) TYPE: nucleic acid
1006             (C) STRANDEDNESS: single
1007             (D) TOPOLOGY: linear
1008         (ii) MOLECULE TYPE: Primer DNA (genomic) human
1009         (ix) FEATURE:
1010             (A) NAME/KEY: AIPL1 primer
1011             (B) LOCATION:
1012             (D) OTHER INFORMATION:
1013         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:
E--> 1015         5'-ACGCAGAGGTGTGGAATG-3'                18
1017 (2) INFORMATION FOR SEQ ID NO: 49:
1018         (i) SEQUENCE CHARACTERISTICS:
1019             (A) LENGTH: 19 bases
1020             (B) TYPE: nucleic acid
1021             (C) STRANDEDNESS: single
1022             (D) TOPOLOGY: linear
1023         (ii) MOLECULE TYPE: Primer DNA (genomic) human
1024         (ix) FEATURE:
1025             (A) NAME/KEY: AIPL1 Primer
1026             (B) LOCATION:
1027             (D) OTHER INFORMATION:
1028         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:
E--> 1030         5'-AAAAAGTGACACCACGATC-3'                19
1112 (2) INFORMATION FOR SEQ ID NO: 55:
1113         (i) SEQUENCE CHARACTERISTICS:
1114             (A) LENGTH: 6689 bases 35 shown
1115             (B) TYPE: nucleic acid
1116             (C) STRANDEDNESS: single
1117             (D) TOPOLOGY: linear
1118         (ii) MOLECULE TYPE: cDNA
1119         (ix) FEATURE:
1120             (A) NAME/KEY: AIPL1 gene exon/intron Acceptor
1121 splice site
1122             (B) LOCATION:
1123             (D) OTHER INFORMATION:
1124         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:
E--> 1126         CACTGACCTGCAGCTCTGGGGCCAGGTTGATGCCC        35
1193 (2) INFORMATION FOR SEQ ID NO: 60:
1194         (i) SEQUENCE CHARACTERISTICS:
1195             (A) LENGTH: 18 bases
1196             (B) TYPE: nucleic acid
1197             (C) STRANDEDNESS: single

```

grow bases into 10's

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/765,061A

DATE: 08/06/2001

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

1198         (D) TOPOLOGY: linear
W--> 1199     (ii) MOLECULE TYPE: DNA Primer
1200         (ix) FEATURE:
1201             (A) NAME/KEY: AIPL1 gene Exon 1 Primer
1202             (B) LOCATION: 240
1203             (D) OTHER INFORMATION:
1204         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:
E--> 1206         5'-GGACACCTCCCTTTCTCC-3'
1208 (2) INFORMATION FOR SEQ ID NO: 61:
1209     (i) SEQUENCE CHARACTERISTICS:
1210         (A) LENGTH: 18 bases
1211         (B) TYPE: nucleic acid
1212         (C) STRANDEDNESS: single
1213         (D) TOPOLOGY: linear
W--> 1214     (ii) MOLECULE TYPE: DNA Primer
1215         (ix) FEATURE:
1216             (A) NAME/KEY: AIPL1 gene Exon 1 Primer
1217             (B) LOCATION: 240
1218             (D) OTHER INFORMATION:
1219         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:
E--> 1221         5'-GCTGGGGCTGCCTGGCTG-3'
1223 (2) INFORMATION FOR SEQ ID NO: 62:
1224     (i) SEQUENCE CHARACTERISTICS:
1225         (A) LENGTH: 20 bases
1226         (B) TYPE: nucleic acid
1227         (C) STRANDEDNESS: single
1228         (D) TOPOLOGY: linear
W--> 1229     (ii) MOLECULE TYPE: DNA Primer
1230         (ix) FEATURE:
1231             (A) NAME/KEY: AIPL1 gene Exon 2 Primer
1232             (B) LOCATION: 297
1233             (D) OTHER INFORMATION:
1234         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:
E--> 1236         5'-GGGCCTTGAACAGTGTGTCT-3'
1238 (2) INFORMATION FOR SEQ ID NO: 63:
1239     (i) SEQUENCE CHARACTERISTICS:
1240         (A) LENGTH: 19 bases
1241         (B) TYPE: nucleic acid
1242         (C) STRANDEDNESS: single
1243         (D) TOPOLOGY: linear
W--> 1244     (ii) MOLECULE TYPE: DNA Primer
1245         (ix) FEATURE:
1246             (A) NAME/KEY: AIPL1 gene Exon 2 Primer
1247             (B) LOCATION: 297
1248             (D) OTHER INFORMATION:
1249         (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:
E--> 1251         5'-TTTCCCGAAACACAGCAGC-3'
1253 (2) INFORMATION FOR SEQ ID NO: 64:
1254     (i) SEQUENCE CHARACTERISTICS:

```

18

18

20

19

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/765,061A

DATE: 08/06/2001

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

1255 (A) LENGTH: 18 bases
 1256 (B) TYPE: nucleic acid
 1257 (C) STRANDEDNESS: single
 1258 (D) TOPOLOGY: linear

W--> 1259 (ii) MOLECULE TYPE: DNA Primer
 1260 (ix) FEATURE:
 1261 (A) NAME/KEY: AIPL1 gene Exon 3 Primer
 1262 (B) LOCATION: 364
 1263 (D) OTHER INFORMATION:
 1264 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

E--> 1266 5'-AGTGGAGGAGCAGGATTC-3' 18

1373 (2) INFORMATION FOR SEQ ID NO: 72:
 1374 (i) SEQUENCE CHARACTERISTICS: 384
 1375 (A) LENGTH: 383 amino acids
 1376 (B) TYPE: amino acid
 1377 (D) TOPOLOGY: linear
 1378 (ii) MOLECULE TYPE: protein
 1379 (ix) FEATURE:
 1380 (A) NAME/KEY: Human Aipl1
 1381 (B) LOCATION:
 1382 (D) OTHER INFORMATION:
 1383 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

1385	Met	Asp	Ala	Ala	Leu	Leu	Leu	Asn	Val	Glu	Gly	Val	Lys	Lys	Thr
1386	1				5					10					15
1387	Ile	Leu	His	Gly	Gly	Thr	Gly	Glu	Leu	Pro	Asn	Phe	Ile	Thr	Gly
1388					20					25					30
1389	Ser	Arg	Val	Ile	Phe	His	Phe	Arg	Thr	Met	Lys	Cys	Asp	Glu	Glu
1390					35					40					45
1391	Arg	Thr	Val	Ile	Asp	Ser	Arg	Gln	Val	Gly	Gln	Pro	Met	His	
1392					50					55					60
1393	Ile	Ile	Ile	Gly	Asn	Met	Phe	Lys	Leu	Glu	Val	Trp	Glu	Ile	Leu
1394					65					70					75
1395	Leu	Thr	Ser	Met	Arg	Val	His	Glu	Val	Ala	Glu	Phe	Trp	Cys	Asp
1396					80					85					90
1397	Thr	Ile	His	Thr	Gly	Val	Tyr	Pro	Ile	Leu	Ser	Arg	Ser	Leu	Arg
1398					95					100					105
1399	Gln	Met	Ala	Gln	Gly	Lys	Asp	Pro	Thr	Glu	Trp	His	Val	His	Thr
1400					110					115					120
1401	Cys	Gly	Leu	Ala	Asn	Met	Phe	Ala	Tyr	His	Thr	Leu	Gly	Tyr	Glu
1402					125					130					135
1403	Asp	Leu	Asp	Glu	Leu	Gln	Lys	Glu	Pro	Gln	Pro	Leu	Val	Phe	Val
1404					140					145					150
1405	Ile	Glu	Leu	Leu	Gln	Val	Asp	Ala	Pro	Ser	Asp	Tyr	Gln	Arg	Glu
1406					155					160					165
1407	Thr	Trp	Asn	Leu	Ser	Asn	His	Glu	Lys	Met	Lys	Ala	Val	Pro	Val
1408					170					175					180
1409	Leu	His	Gly	Glu	Gly	Asn	Arg	Leu	Phe	Lys	Leu	Gly	Arg	Tyr	Glu
1410					185					190					195
1411	Glu	Ala	Ser	Ser	Lys	Tyr	Gln	Glu	Ala	Ile	Ile	Cys	Leu	Arg	Asn

The types of errors shown exist throughout the Sequence Listing. Please check subsequent sequences for similar errors.

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

```

1412                200                205                210
1413 Leu Gln Thr Lys Glu Lys Pro Trp Glu Val Gln Trp Leu Lys Leu
1414                215                220                225
1415 Glu Lys Met Ile Asn Thr Leu Ile Leu Asn Tyr Cys Gln Cys Leu
1416                230                235                240
1417 Leu Lys Lys Glu Glu Tyr Tyr Glu Val Leu Glu His Thr Ser Asp
1418                245                250                255
1419 Ile Leu Arg His His Pro Gly Ile Val Lys Ala Tyr Tyr Val Arg
1420                260                265                270
1421 Ala Arg Ala His Ala Glu Val Trp Asn Glu Ala Glu Ala Lys Ala
1422                275                280                285
1423 Asp Leu Gln Lys Val Leu Glu Leu Glu Pro Ser Met Gln Lys Ala
1424                290                295                300
1425 Val Arg Arg Glu Leu Arg Leu Leu Glu Asn Arg Met Ala Glu Lys
1426                305                310                315
E--> 1427 Gln Glu Glu Glu Arg Leu (Xxx) Cys Arg Asn Met Leu Ser Gln Gly
1428                320                325                330
1429 Ala Thr Gln Pro Pro Ala Glu Pro Pro Thr Glu Pro Pro Ala Gln
1430                335                340                345
1431 Ser Ser Thr Glu Pro Pro Ala Glu Pro Pro Thr Ala Pro Ser Ala
1432                350                355                360
1433 Glu Leu Ser Ala Gly Pro Pro Ala Glu Pro Ala Thr Glu Pro Pro
1434                365                370                375
1435 Pro Ser Pro Gly His Ser Leu Gln His
E--> 1436                380                383
1438 (2) INFORMATION FOR SEQ ID NO: 73:
1439 (i) SEQUENCE CHARACTERISTICS:
1440 (A) LENGTH: 369 amino acids
1441 (B) TYPE: amino acid
1442 (D) TOPOLOGY: linear
1443 (ii) MOLECULE TYPE: protein
1444 (ix) FEATURE:
1445 (A) NAME/KEY: Chimpanzee Aipl1
1446 (B) LOCATION:
1447 (D) OTHER INFORMATION:
1448 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:
1450 Met Asp Ala Ala Leu Leu Leu Asn Val Glu Gly Val Lys Lys Thr
1451 1 5 10 15
1452 Ile Leu His Gly Gly Thr Gly Glu Leu Pro Asn Phe Ile Thr Gly
1453 20 25 30
1454 Ser Arg Val Ile Phe His Phe Arg Thr Met Lys Cys Asp Glu Glu
1455 35 40 45
1456 Arg Thr Val Ile Asp Asp Ser Arg Gln Val Gly Gln Pro Met His
1457 50 55 60
1458 Ile Ile Ile Gly Asn Met Phe Lys Leu Glu Val Trp Glu Ile Leu
1459 65 70 75
1460 Leu Thr Ser Met Arg Val His Glu Val Ala Glu Phe Trp Cys Asp
1461 80 85 90
1462 Thr Ile His Thr Gly Val Tyr Pro Ile Leu Ser Arg Ser Leu Arg

```

RAW SEQUENCE LISTING

DATE: 08/06/2001

PATENT APPLICATION: US/09/765,061A

TIME: 09:19:35

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

1463		95		100		105
1464	Gln Met Ala Gln Gly Lys Asp Pro Thr		Glu Trp His Val His Thr			
1465		110		115		120
1466	Cys Gly Leu Ala Asn Met Phe Ala Tyr		His Thr Leu Gly Tyr Glu			
1467		125		130		135
1468	Asp Leu Asp Glu Leu Gln Lys Glu Pro		Gln Pro Leu Val Phe Val			
1469		140		145		150
1470	Ile Glu Leu Leu Gln Val Asp Ala Pro		Ser Asp Tyr Gln Arg Glu			
1471		155		160		165
1472	Thr Trp Asn Leu Ser Asn His Glu Lys		Met Lys Ala Val Pro Val			
1473		170		175		180
1474	Leu His Gly Glu Gly Asn Arg Leu Phe		Lys Leu Gly Arg Tyr Glu			
1475		185		190		195
1476	Glu Ala Ser Ser Lys Tyr Gln Glu Ala		Ile Ile Cys Leu Arg Asn			
1477		200		205		210
1478	Leu Gln Thr Lys Glu Lys Pro Trp Glu		Val Gln Trp Leu Lys Leu			
1479		215		220		225
1480	Glu Lys Met Ile Asn Thr Leu Ile Leu		Asn Tyr Cys Gln Cys Leu			
1481		230		235		240
1482	Leu Lys Lys Glu Glu Tyr Tyr Glu Val		Leu Glu His Thr Ser Asp			
1483		245		250		255
1484	Ile Leu Arg His His Pro Gly Ile Val		Lys Ala Tyr Tyr Val Arg			
1485		260		265		270
1486	Ala Arg Ala His Ala Glu Val Trp Asn		Glu Ala Glu Ala Lys Ala			
1487		275		280		285
1488	Asp Leu Arg Lys Val Leu Glu Leu Glu		Pro Ser Met Gln Lys Ala			
1489		290		295		300
1490	Val Arg Arg Glu Leu Arg Leu Leu Glu		Asn Arg Met Ala Glu Lys			
1491		305		310		315
1492	Gln Glu Glu Glu Arg Leu Arg Cys Arg		Asn Met Leu Ser Gln Gly			
1493		320		325		330
1494	Ala Thr Gln Pro Pro Ala Glu Pro Pro		Thr Glu Pro Pro Ala Gln			
1495		335		340		345
1496	Ser Ser Thr Glu Pro Pro Ala Glu Pro		Pro Pro Ala Pro Ser Ala			
1497		350		355		360
1498	Glu Leu Ser Ala Gly Pro Pro Ala Glu		Thr Ala Thr Glu Pro Pro			
1499		365		370		375
1500	Pro Ser Pro Gly His Ser Leu Gln His					

E--> 1501

~~365~~~~369~~

380

see attached

(3) Computer: Apple Macintosh;
 (i) Operating System: Macintosh;
 (ii) Macintosh File Type: text with line termination
 (iii) Line Terminator: Pre-defined by text type file;
 (iv) Pagination: Pre-defined by text type file;
 (v) End-of-file: Pre-defined by text type file;
 (vi) Media: (A) Diskett—3.50 Inch, 400 Kb storage;
 (B) Diskette—3.50 Inch, 800 Kb storage;
 (C) Diskette—3.50 Inch, 1.4 Mb storage;
 (vii) Print Command: Use PRINT command from any Macintosh Application that processes text files, such as MacWrite or Teach Text;
 (4) Magnetic tape: 0.5 Inch, up to 2400 feet;
 (i) Density: 1600 or 6250 bits per Inch, 9 track;
 (ii) Format: raw, unblocked;
 (iii) Line Terminator: ASCII Carriage Return plus optional ASCII Line Feed;
 (iv) Pagination: ASCII Form Feed or Series of Line Terminators;
 (v) Print Command (Unix shell version given here as sample response—mt/dev/rmt0; lpr/dev/rmt0);
 (g) Computer readable forms that are submitted to the Office will not be returned to the applicant.
 (h) All computer readable forms shall have a label permanently affixed thereto on which has been hand-printed or typed, a description of the format of the computer readable form as well as the name of the applicant, the title of the invention, the date on which the data were recorded on the computer readable form and the name and type of computer and operating system which generated the files on the computer readable form. If all of this information cannot be printed on a label affixed to the computer readable form, by reason of size or otherwise, the label shall include the name of the applicant and the title of the invention and a reference number, and the additional information may be provided on a container for the computer readable form with the name of the applicant, the title of the invention, the reference number and the additional information affixed to the container. If the computer readable form is submitted after the date of filing

under 35 U.S.C. 111, after the date of entry in the national stage under 35 U.S.C. 371 or after the time of filing, in the United States Receiving Office, an international application under the PCT, the labels mentioned herein must also include the date of the application and the application number, including series code and serial number.

§ 1.825 Amendments to or replacement of sequence listing and computer readable copy thereof.

(a) Any amendment to the paper copy of the "Sequence Listing" (§ 1.821(c)) must be made by the submission of substitute sheets. Amendments must be accompanied by a statement that indicates support for the amendment in the application, as filed, and a statement that the substitute sheets include no new matter. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(b) Any amendment to the paper copy of the "Sequence Listing," in accordance with paragraph (a) of this section, must be accompanied by a substitute copy of the computer readable form (§ 1.821(e)) including all previously submitted data with the amendment incorporated therein, accompanied by a statement that the copy in computer readable form is the same as the substitute copy of the "Sequence Listing." Such a statement must be a verified statement if made by a person not registered to practice before the Office.

(c) Any appropriate amendments to the "Sequence Listing" in a patent, e.g., by reason of reissue or certificate of correction, must comply with the requirements of paragraphs (a) and (b) of this section.

(d) If, upon receipt, the computer readable form is found to be damaged or unreadable, applicant must provide, within such time as set by the Commissioner, a substitute copy of the data in computer readable form accompanied by a statement that the substitute data is identical to that originally filed. Such a statement must be a verified statement if made by a person not registered to practice before the Office.

Appendix A—Sample Sequence Listing

(1) GENERAL INFORMATION:

(I) APPLICANT: Doe, Joan X. Doe, John Q
 (II) TITLE OF INVENTION: Isolation and Characterization of a Gene Encoding a Protease from *Paramecium* sp.
 (III) NUMBER OF SEQUENCES: 2
 (IV) CORRESPONDENCE ADDRESS:
 (A) ADDRESSEE: Smith and Jones
 (B) STREET: 123 Main Street
 (C) CITY: Smalltown
 (D) STATE: Anystate
 (E) COUNTRY: USA
 (F) ZIP: 12345
 (V) COMPUTER READABLE FORM:
 (A) MEDIUM TYPE: Diskette, 3.50 Inch, 800 Kb storage
 (B) COMPUTER: Apple Macintosh
 (C) OPERATING SYSTEM: McIntosh 5.0
 (D) SOFTWARE: MacWrite
 (VI) CURRENT APPLICATION DATA:
 (A) APPLICATION NUMBER: 09/999,999
 (B) FILING DATE: 28-FEB-1989
 (C) CLASSIFICATION: 999/99
 (VII) PRIOR APPLICATION DATA:
 (A) APPLICATION NUMBER: PCT/US88/099999
 (B) FILING DATE: 01-MAR-1988
 (VIII) ATTORNEY/AGENT INFORMATION:
 (A) NAME: Smith, John A.
 (B) REGISTRATION NUMBER: 00001
 (C) REFERENCE/DOCKET NUMBER: 01-0001
 (IX) TELECOMMUNICATION INFORMATION:
 (A) TELEPHONE: (909) 999-0001
 (B) TELEFAX: (909) 999-0002
 (2) INFORMATION FOR SEQ ID NO: 1:
 (I) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 954 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (II) MOLECULE TYPE: genomic DNA
 (III) HYPOTHETICAL: yes
 (IV) ANTI-SENSE: no
 (VI) ORIGINAL SOURCE:
 (A) ORGANISM: *Paramecium* sp.
 (C) INDIVIDUAL/ISOLATE: XYZ2
 (G) CELL TYPE: unicellular organism
 (VII) IMMEDIATE SOURCE:
 (A) LIBRARY: genomic
 (B) CLONE: Para-XYZ2/36
 (X) PUBLICATION INFORMATION:
 (A) AUTHORS: Doe, Joan X. Doe, John Q
 (B) TITLE: Isolation and Characterization of a Gene Encoding a-Protease from *Paramecium* sp.
 (C) JOURNAL: Fictional Genes
 (D) VOLUME: 1
 (E) ISSUE: 1
 (F) PAGES: 1-20
 (G) DATE: 02-MAR-1988
 (K) RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 954

BILLING CODE 3510-10-M

ATCGGGATAG	TACTGGTCAA	GACCGGTGGA	CACCGGTAA	CCCCGGTTAA	GTACCGGTTA	60
TAGGCCATTT	CAGGCCAAAT	GTGCCCAACT	ACGCCAATTG	TTTTGCCAAC	GGCCAACGTT	120
ACGTTCGTAC	GCACGTATGT	ACCTAGGTAC	TTACGGACGT	GACTACGGAC	ACTTCCGTAC	180
GTACGTACGT	TTACGTACCC	ATCCCAACGT	AACCACAGTG	TGGTCGCAGT	GTCCCAGTGT	240
ACACAGACTG	CCAGACATTC	TTCACAGACA	CCCC	ATG	ACA	295
				CCT	GAA	
				CGT	CTC	
				Met	Thr	
				Pro	Pro	
				Glu	Arg	
				Leu		

TTC CTC CCA AGG GTG TGT GGC ACC ACC CTA CAC CTC CTC CTT CTG GGG 343
Phe Leu Pro Arg Val Cys Gly Thr Thr Leu His Leu Leu Leu Leu Gly
-25 -20 -15

CTG CTG CTG GTT CTG CTG CCT GGG GCC CAT GTGAGGCAGC AGGAGAATGG 393
Leu Leu Leu Val Leu Leu Pro Gly Ala His
-10 -5

GGTGGCTCAG CCAAACCTTG AGCCCTAGAG CCCCCCTCAA CTCTGTTCTC CTAG GGG 450
Gly

CTC ATG CAT CTT GCC CAC AGC AAC CTC AAA CCT GCT GCT CAC CTC ATT 498
Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His Leu Ile
1 5 10 15

GTAAACATCC ACCTGACCTC CCAGACATGT CCCCACCAGC TCTCCTCCTA CCCCTGCCTC 558

AGGAACCCAA GCATCCACCC CTCTCCCCCA ACTTCCCCCA CGCTAAAAAA AACAGAGGGA 618

GGCCACTCCT ATGCCTCCCC CTGCCATCCC CCAGGAACTC AGTTGTTTCTAG TGCCCACTTC 678

TAC CCC AGC AAG CAG AAC TCA CTG CTC TGG AGA GCA AAC ACG GAC CGT 726
Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr Asp Arg
20 25 30

GCC TTC CTC CAG GAT GGT TTC TCC TTG AGC AAC AAT TCT CTC CTG GTC 774
Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu Leu Val
35 40 45

TAGAAAAAAT AATTGATTTC AAGACCTTCT CCCCATTTCTG CCTCCATTCT GACCATTTC 834

GGGGTCGTCA CCACCTCTCC TTTGGCCATT CCAACAGCTC AAGTCTTCCC TGATCAAGTC 894

ACCGGAGCTT TCAAAGAAGG AATTCTAGGC ATCCCAGGGG ACCCACACCT CCCTGAACCA 954

(2) INFORMATION FOR SEQ ID NO: 2:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 82 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: protein
(ix) FEATURE:
(A) NAME/KEY: signal sequence
(B) LOCATION: -34 to -1

(C) IDENTIFICATION METHOD: similarity
to other signal sequences, hydrophobic
(D) OTHER INFORMATION: expresses
protease
(x) PUBLICATION INFORMATION:
(A) AUTHORS: Doe, Joan X. Doe, John Q
(B) TITLE: Isolation and Characterization
of a Gene Encoding a Protease from
Paramecium sp.

(C) JOURNAL: Fictional Genes
(D) VOLUME: 1
(E) ISSUE: 1
(F) PAGES: 1-20
(G) DATE: 02-MAR-1988
(K) RELEVANT RESIDUES IN SEQ ID NO:
2: FROM -34 TO 48
BILLING CODE 2610-10-M

Here's where sequence 2 starts (after
the sequence data of SEQ ID NO: 1:)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Pro Pro Glu Arg Leu Phe Leu Pro Arg Val Cys Gly Thr Thr
-30 -25 -20

Leu His Leu Leu Leu Leu Gly Leu Leu Leu Val Leu Leu Pro Gly Ala
-15 -10 -5

His Gly Leu Met His Leu Ala His Ser Asn Leu Lys Pro Ala Ala His
1 5 10

Leu Ile Tyr Pro Ser Lys Gln Asn Ser Leu Leu Trp Arg Ala Asn Thr
15 20 25 30

Asp Arg Ala Phe Leu Gln Asp Gly Phe Ser Leu Ser Asn Asn Ser Leu
35 40 45

Leu Val

BILLING CODE 3510-16-C

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/765,061A

DATE: 08/06/2001

TIME: 09:19:36

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

L:0 M:200 E: Mandatory Header Field missing, [(i) APPLICANT:] of (1)
 L:0 M:200 E: Mandatory Header Field missing, [(ii) TITLE OF INVENTION:] of (1)
 L:0 M:200 E: Mandatory Header Field missing, [(A) ADDRESSEE:] of (1)(iv)
 L:0 M:200 E: Mandatory Header Field missing, [(B) STREET:] of (1)(iv)
 L:0 M:200 E: Mandatory Header Field missing, [(C) CITY:] of (1)(iv)
 L:0 M:249 C: Inserted Mandatory Field, [(vi) CURRENT APPLICATION DATA:]
 L:0 M:249 C: Inserted Mandatory Field, [(A) APPLICATION NUMBER:]
 L:0 M:249 C: Inserted Mandatory Field, [(B) FILING DATE:]
 L:18 M:111 C: (47) String data converted to upper case,
 M:111 Repeated in SeqNo=1
 L:129 M:254 E: No. of Bases conflict, Input:6749 Counted:6689 SEQ:1
 L:129 M:204 E: No. of Bases differ, LENGTH:Input:6749 Counted:6689 SEQ:1
 L:145 M:111 C: (47) String data converted to upper case,
 M:111 Repeated in SeqNo=2
 L:179 M:111 C: (47) String data converted to upper case,
 M:111 Repeated in SeqNo=3
 L:214 M:111 C: (47) String data converted to upper case,
 M:111 Repeated in SeqNo=4
 L:248 M:111 C: (47) String data converted to upper case,
 M:111 Repeated in SeqNo=5
 L:279 M:111 C: (47) String data converted to upper case,
 M:111 Repeated in SeqNo=6
 L:312 M:111 C: (47) String data converted to upper case,
 M:111 Repeated in SeqNo=7
 L:347 M:111 C: (47) String data converted to upper case,
 M:111 Repeated in SeqNo=8
 L:365 M:254 E: No. of Bases conflict, Input:1129 Counted:1119 SEQ:8
 L:365 M:204 E: No. of Bases differ, LENGTH:Input:1129 Counted:1119 SEQ:8
 L:380 M:111 C: (47) String data converted to upper case,
 L:396 M:111 C: (47) String data converted to upper case,
 L:397 M:254 E: No. of Bases conflict, Input:0 Counted:15 SEQ:10
 L:397 M:320 E: (1) Wrong Nucleic Acid Designator, 1
 L:397 M:204 E: No. of Bases differ, LENGTH:Input:15 Counted:16 SEQ:10
 L:412 M:111 C: (47) String data converted to upper case,
 L:428 M:111 C: (47) String data converted to upper case,
 L:444 M:111 C: (47) String data converted to upper case,
 L:460 M:111 C: (47) String data converted to upper case,
 L:461 M:254 E: No. of Bases conflict, Input:0 Counted:15 SEQ:14
 L:461 M:320 E: (1) Wrong Nucleic Acid Designator, 1
 L:461 M:204 E: No. of Bases differ, LENGTH:Input:15 Counted:16 SEQ:14
 L:476 M:111 C: (47) String data converted to upper case,
 L:492 M:111 C: (47) String data converted to upper case,
 L:493 M:254 E: No. of Bases conflict, Input:0 Counted:15 SEQ:16
 L:493 M:320 E: (1) Wrong Nucleic Acid Designator, 1
 L:493 M:204 E: No. of Bases differ, LENGTH:Input:15 Counted:16 SEQ:16
 L:508 M:111 C: (47) String data converted to upper case,
 L:509 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:509 M:333 E: Wrong sequence grouping, Amino acids not in groups!

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/765,061A

DATE: 08/06/2001

TIME: 09:19:36

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

L:524 M:111 C: (47) String data converted to upper case,
 L:540 M:111 C: (47) String data converted to upper case,
 L:557 M:111 C: (47) String data converted to upper case,
 L:558 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:558 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:573 M:111 C: (47) String data converted to upper case,
 L:589 M:111 C: (47) String data converted to upper case,
 L:590 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:590 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:605 M:111 C: (47) String data converted to upper case,
 L:606 M:254 E: No. of Bases conflict, Input:0 Counted:15 SEQ:23
 L:606 M:320 E: (1) Wrong Nucleic Acid Designator, 1 ✓
 L:606 M:204 E: No. of Bases differ, LENGTH:Input:15 Counted:16 SEQ:23
 L:621 M:111 C: (47) String data converted to upper case,
 L:622 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:622 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:637 M:111 C: (47) String data converted to upper case,
 L:638 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:638 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:654 M:111 C: (47) String data converted to upper case,
 L:655 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:655 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:671 M:111 C: (47) String data converted to upper case,
 L:672 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:672 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:688 M:111 C: (47) String data converted to upper case,
 L:689 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:689 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:705 M:111 C: (47) String data converted to upper case,
 L:706 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:706 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:722 M:111 C: (47) String data converted to upper case,
 L:723 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:723 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:739 M:111 C: (47) String data converted to upper case,
 L:740 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:740 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:756 M:111 C: (47) String data converted to upper case,
 L:757 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:757 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:773 M:111 C: (47) String data converted to upper case,
 L:774 M:334 W: (2) Invalid Amino Acid in Coding Region, NUMBER OF INVALID KEYS:2
 L:774 M:333 E: Wrong sequence grouping, Amino acids not in groups! ✓
 L:790 M:111 C: (47) String data converted to upper case,
 L:807 M:111 C: (47) String data converted to upper case,
 L:824 M:111 C: (47) String data converted to upper case,
 L:841 M:111 C: (47) String data converted to upper case,
 L:858 M:111 C: (47) String data converted to upper case,
 L:875 M:111 C: (47) String data converted to upper case,

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/765,061A

DATE: 08/06/2001

TIME: 09:19:36

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

L:892 M:111 C: (47) String data converted to upper case,
 L:909 M:111 C: (47) String data converted to upper case,
 L:918 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=42
 L:925 M:254 E: No. of Bases conflict, Input:20 Counted:20 SEQ:42
 L:925 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 M:111 Repeated in SeqNo=42
 L:925 M:204 E: No. of Bases differ, LENGTH:Input:20 Counted:24 SEQ:42
 L:933 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=43
 L:940 M:254 E: No. of Bases conflict, Input:19 Counted:19 SEQ:43
 L:940 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:940 M:204 E: No. of Bases differ, LENGTH:Input:19 Counted:23 SEQ:43
 L:948 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=44
 L:955 M:254 E: No. of Bases conflict, Input:17 Counted:17 SEQ:44
 L:955 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:955 M:204 E: No. of Bases differ, LENGTH:Input:17 Counted:21 SEQ:44
 L:970 M:254 E: No. of Bases conflict, Input:18 Counted:18 SEQ:45
 L:970 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:970 M:204 E: No. of Bases differ, LENGTH:Input:18 Counted:22 SEQ:45
 L:985 M:254 E: No. of Bases conflict, Input:20 Counted:20 SEQ:46
 L:985 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:985 M:204 E: No. of Bases differ, LENGTH:Input:20 Counted:24 SEQ:46
 L:1000 M:254 E: No. of Bases conflict, Input:20 Counted:20 SEQ:47
 L:1000 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:1000 M:204 E: No. of Bases differ, LENGTH:Input:20 Counted:24 SEQ:47
 L:1015 M:254 E: No. of Bases conflict, Input:18 Counted:18 SEQ:48
 L:1015 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:1015 M:204 E: No. of Bases differ, LENGTH:Input:18 Counted:22 SEQ:48
 L:1030 M:254 E: No. of Bases conflict, Input:19 Counted:19 SEQ:49
 L:1030 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:1030 M:204 E: No. of Bases differ, LENGTH:Input:19 Counted:23 SEQ:49
 L:1126 M:333 E: Wrong sequence grouping, Nucleotides not in groups!
 L:1126 M:204 E: No. of Bases differ, LENGTH:Input:6689 Counted:35 SEQ:55
 L:1199 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=60
 L:1206 M:254 E: No. of Bases conflict, Input:18 Counted:18 SEQ:60
 L:1206 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:1206 M:204 E: No. of Bases differ, LENGTH:Input:18 Counted:22 SEQ:60
 L:1214 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=61
 L:1221 M:254 E: No. of Bases conflict, Input:18 Counted:18 SEQ:61
 L:1221 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:1221 M:204 E: No. of Bases differ, LENGTH:Input:18 Counted:22 SEQ:61
 L:1229 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=62
 L:1236 M:254 E: No. of Bases conflict, Input:20 Counted:20 SEQ:62
 L:1236 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:1236 M:204 E: No. of Bases differ, LENGTH:Input:20 Counted:24 SEQ:62
 L:1244 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=63
 L:1251 M:254 E: No. of Bases conflict, Input:19 Counted:19 SEQ:63
 L:1251 M:320 E: (1) Wrong Nucleic Acid Designator, 6
 L:1251 M:204 E: No. of Bases differ, LENGTH:Input:19 Counted:23 SEQ:63
 L:1259 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=64

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/765,061A

DATE: 08/06/2001

TIME: 09:19:36

Input Set : A:\converted sequences v2.txt

Output Set: N:\CRF3\08062001\I765061A.raw

L:1266 M:254 E: No. of Bases conflict, Input:18 Counted:18 SEQ:64
L:1266 M:320 E: (1) Wrong Nucleic Acid Designator, 6
L:1266 M:204 E: No. of Bases differ, LENGTH:Input:18 Counted:22 SEQ:64
L:1274 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=65
L:1281 M:254 E: No. of Bases conflict, Input:20 Counted:20 SEQ:65
L:1281 M:320 E: (1) Wrong Nucleic Acid Designator, 6
L:1281 M:204 E: No. of Bases differ, LENGTH:Input:20 Counted:24 SEQ:65
L:1289 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=66
L:1296 M:254 E: No. of Bases conflict, Input:18 Counted:18 SEQ:66
L:1296 M:320 E: (1) Wrong Nucleic Acid Designator, 6
L:1296 M:204 E: No. of Bases differ, LENGTH:Input:18 Counted:22 SEQ:66
L:1304 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=67
L:1311 M:254 E: No. of Bases conflict, Input:17 Counted:17 SEQ:67
L:1311 M:320 E: (1) Wrong Nucleic Acid Designator, 6
L:1311 M:204 E: No. of Bases differ, LENGTH:Input:17 Counted:21 SEQ:67
L:1319 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=68
L:1326 M:254 E: No. of Bases conflict, Input:19 Counted:19 SEQ:68
L:1326 M:320 E: (1) Wrong Nucleic Acid Designator, 6
L:1326 M:204 E: No. of Bases differ, LENGTH:Input:19 Counted:23 SEQ:68
L:1334 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=69
L:1341 M:254 E: No. of Bases conflict, Input:18 Counted:18 SEQ:69
L:1341 M:320 E: (1) Wrong Nucleic Acid Designator, 6
L:1341 M:204 E: No. of Bases differ, LENGTH:Input:18 Counted:22 SEQ:69
L:1349 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=70
L:1356 M:254 E: No. of Bases conflict, Input:18 Counted:18 SEQ:70
L:1356 M:320 E: (1) Wrong Nucleic Acid Designator, 6
L:1356 M:204 E: No. of Bases differ, LENGTH:Input:18 Counted:22 SEQ:70
L:1364 M:246 W: Invalid value of Alpha Sequence Header Field, [MOLECULE TYPE:], SeqNo=71
L:1371 M:254 E: No. of Bases conflict, Input:19 Counted:19 SEQ:71
L:1371 M:320 E: (1) Wrong Nucleic Acid Designator, 6
L:1371 M:204 E: No. of Bases differ, LENGTH:Input:19 Counted:23 SEQ:71
L:1427 M:330 E: (2) Invalid Amino Acid Designator, 1
L:1436 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:72
L:1436 M:203 E: No. of Seq. differs, LENGTH:Input:383 Found:384 SEQ:72
L:1501 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:73
L:1501 M:203 E: No. of Seq. differs, LENGTH:Input:369 Found:384 SEQ:73
L:3 M:203 E: No. of Seq. differs, : Input 1, Counted 78